AFGL-TR-81-0056

CHEBYSHEV POLYNOMIAL APPROXIMATION
IN THE OPAQUE PROGRAM

P. Tsipouras
T. Costello

Bedford Research Associates
2 DeAngelo Drive
Bedford, Massachusetts 01730

Scientific Report No. 4

30 January 1981

DTIC
ELECTE
SEP 17 1981

D

81 9 17 029

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
AFGL-TR-81-0056	AD-A104317	

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
Chebyshev Polynomial Approximation in the OPAQUE program.	Scientific Report No. 4
	6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
P. Tsipouras * T. Costello	F19628-79-C-0163

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Bedford Research Associates 2 DeAngelo Drive Bedford, MA. 01730	62101 9993XXXX

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Air Force Geophysics Laboratory Hanscom AFB, Massachusetts 01731 Monitor/Paul Tsipouras/SUWA	30 January 1981
	13. NUMBER OF PAGES
	14

14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)	15. SECURITY CLASS. (of this report)
	Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

SCIENTIFIC-4

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, If different from Report)

Accession For

NTIS GRA&I ☒
DTIC TAB ☐
Unannounced ☐
Justification

By
Distribution/
Availability Codes

| Dist | Avail and/or Special |
| A | |

18. SUPPLEMENTARY NOTES

* Air Force Geophysics Laboratories
Analysis and Simulation Branch (SUWA)
Hanscom AFB, MA. 01731

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Chebyshev, Aerosol, Lognormal.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The use of expansions in Chebyshev polynomials is investigated for determining the size distribution function of aerosol particles. Three different models of size distribution in aerosols, Jung's, a unimodel lognormal model, and the haze L model are considered in this context, and the advantages of the techniques are indicated.

DD FORM, JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

BLANK PAGE

1. INTRODUCTION

The atmospheric environment is a decisive influence on many electro-optical systems. In most cases, the atmospheric medium constitutes a limiting factor for the propagation of electro-magnetic signals. In other cases, the optical properties of this medium render systems infeasible. Thus to support electro-optics system planners, designers and users with probability of occurrence information on earth surface propagation conditions, it is desirable to have statistics of specific propagation parameters available.

The natural aerosol is a basic sonstituent of the atmosphere, having considerable impact generally on optical propagation properties. Its basic parameters are composition, concentration, and distribution.

Though considerable effort has been addressed in the past by many research workers to measuring and understanding the characteristics of natural aerosols, the state-of-the-art must be considered still as quite preliminary. In fact there are neither comprehensive statistics available containing sufficient information for purposes mentioned above, nor unique models which can be used for operational purposes free of ambiguities.

The reason lies in the complexity of aerosol composition, the vast mechanisms of its formation, its complex dynamics, and in its complex way of ageing. One other problem in the past has been the state of data collecting devices which did not easily permit measuring aerosol data over larger size intervals with satisfactory resolution.

The situation is improving now with the advent of new aerosol particle counters.

Typically, aerosol size distribution measurements generate a huge stream of original data. These must be channeled down, for two reasons. First, because of economics, all redundancy in the data should be removed <u>to the largest extent possible</u>. Secondly, there is virtually no way to sensibly work with, analyze, or use in later processing applications, this bulk of original data.

Generally, in a situation like this the benefit of a model consists in providing some kind of mathematical expression through which the data could be greatly compressed. However, there is no such single expression capable of describing size distributions. Junge's formula is referred to very widely in the literature. It is an empirical model in the sense that observation show that it tends to be followed, but not that it must be followed. Yet, for the applications mentioned above, Junge's formula does not seem to be adequate. Other distribution functions are the standard and modified gamma distributions, the normal and the lognormal distributions. Gamma, normal and lognormal distributions have proven to be quite versatile. Generally, they are used for typical background aerosols. Sums of normal and lognormal distributions are used to fit distributions originating from a single source through a single mechanism. These few examples were cited to show that very often the choice of a specific mathematical expression seems to be justified only by the purpose of use. A highly accurate method for fitting aerosol data available involves the use of Spline Functions. However, this approach normally yields a large number of coefficients which do not correspond to physically meaningful parameters.

2. AEROSOL SIZE DISTRIBUTION MEASUREMENTS

2.1 Size Distribution Function

Assume an aerosol is contained in a unit probing volume. The basic method for determining its size distribution function is to count the total number, $N(D)$, of aerosol particles (per unit volume) whose diameter values fall below a certain value, D. The function $N(D)$ is a cumulative function with values between zero and N, the total number of aerosol particles contained within the unit volume. Generally, the number of aerosol particles in a given volume is very large, and their sizes may be assumed to vary continuously. Thus, $N(D)$ approaches a continuous, smooth function as N increases. It is then possible to derive the size distribution function, $n(D)$, by

$$n(D) = \frac{dN(D)}{dD}$$

(2.1)

4

or in integral representation,

$$N(D) = \int_0^D n(D')dD' \qquad (2.2)$$

In practice the determination of N(D) leads to serious problems since it is, for experimental reasons, possible only to measure particles with sizes above a certain, device-dependent, lower limit, say D_L. Thus N(D) is determined up to a constant $N(D_L)$. Since n(D) is a derivative function, this ambiguity is removed.

The aerosol counters employed in this work provide data on the number of particles (per unit volume) for some specified diameter intervals $D_i \leq D \leq D_{i+1}$ with i = 1,2,..., M. If N_i denotes the number of particles for the i-th channel, then because of (2.2) one has .

$$N_i = \int_{D_i}^{D_{i+1}} n(D')\, d\, D' . \qquad (2.3)$$

Accordingly, N_i may be used directly to evaluate an estimate for the size distribution function, n, at the point D_i, that is

$$n(D_i) \approx n_i := \frac{N_i}{D_{i+1} - D_i} \quad ; \quad i = 1,2,...M. \qquad (2.4)$$

3. DATA FIT OBJECTIVES

If data are going to be fitted by some algorithm, two basically different cases are to be distinguished. In one case, the data result from a process, the physical nature of which is understood in the sense that a physical model or theory is available. In this case, data fitting permits the determination of some free model parameters through adjusting the model predictions to the set of measured data. In the other case, there is no physical model at hand

to predict the experimental findings, and data fits in this case are primarily another way of data presentation. If the fit algorithm is chosen carefully and appropriately adapted to the problem, a few, important objectives may be met simultaneously:

A. Data representation:

   The insertion of a smooth curve into the plot of data points is a useful way to assist the understanding of the general trends underlying the data set by inspection.

B. Error smoothing:

   The smooth fit curve is likely to smooth out high frequency contributions within the set of data due to random error. This is especially true if a least-squares-fit algorithm is applied.

C. Data reduction:

   Through data fitting, one can generally reduce the large amount of data to a few functional parameters only. From these, it is possible to retrieve the original data (to the accuracy desired), as well as to calculate new interpolated data.

D. Empirical modelling:

   Generally, it is much easier to correlate any typical condition of the process under study with a typical pattern of the reduced data set, than with the original samples.

The Least Squares Method

Measured data with values v_i ($i = 1,2,...M$) are assumed to be function of a single variable, t. The experimental result is a list of M data pairs (t_i, v_i). The measuring points, t_i, need not be spaced regularly. In the least squares method, a function F from an appropriate collection $\mathcal{F}$ is

6

determined such that

$$\sum_{i=1}^{M} F(t_i) - v_i)^2 = \min. \qquad (4.1)$$

For ease of computation, the function space F is assumed to be formed by all linear combinations of some basis functions, f_k. That is, any $F \in \mathscr{F}$ can be written as

$$F(t) = \sum_{k=0}^{K} b_k f_k(t). \qquad (4.2)$$

The set of f_k must be linearly independent. Though desirable in many cases, the f_k need not necessarily obey some orthogonality condition. In most applications, the function space $\mathscr{F}$ is deliberately restricted by requiring

$$K < M. \qquad (4.3)$$

The free amplitudes, b_k, in (4.2) are determined through (4.1) as solutions of the Gaussian normal equations

$$A \bullet B = C, \qquad (4.4)$$

where A is a positive definite, square symmetric matrix with elements

$$a_{kl} = \sum_{i=1}^{M} f_k(t_i) f_l(t_i), \quad (k, l, = 1, \ldots K), \qquad (4.5)$$

and C is a column vector with elements

$$c_k = \sum_{i=1}^{M} f_k(t_i) \bullet v_i, \quad (k = 1, \ldots K). \qquad (4.6)$$

The vector B of elements b_k is the solution.

The problem encountered in determining solutions of Eq. (4.4) is that the matrix A usually tends to be ill-conditioned. As a result, numerical insta-

7

bilities arise. The problem is reduced somewhat if an appropriate function system, f_k, is used for expansion. Furtheremore, it is advisable not to apply the general Gaussian inversion technique to deduce the solution, B, rather than make use of Cholesky's method, a numerically stable matrix decomposition technique which fully exploits the symmetry and positive definiteness of matrix A.

4.2 Data Transformation

Size distributions of natural aerosols tend to drop off in magnitude over several decades within the diameter range from .1 to 30 m. Therefore, the fit usually is applied not to the original data n_i but rather to their logarithms

$$v_i = \log n_i , \qquad (i = 1, \ldots , M). \qquad (4.7)$$

It should be noted that this is, in fact, a substantial step in the data anlaysis. Indeed, if the fitted curves are going to be employed in any optical propagation calculation, taking the logarithms is a necessary approach. It has been shown that the results of such calculations depend to a very large extent on the number of small aerosol particles as well as on that of larger particles, although the latter usually are much less populated. For the purpose of optical calculations, the "goodness" of fit should be uniform over the entire spectrum of particle diameters considered regardless of their magnitude in concentration. This is quite naturally achieved in using the logarithmic scale since then the least-squares fit results in an overall minimized relative error for the original data.

Typically for the aerosol counters employed in this type of work, the widths of their different size channels generally increase with increasing diameter values, D_1.

8

4.3 Chebyshev Polynomials

The success of the fitting procedure depends critically on the set of basic functions, f_k, to be used for expansion of the data points. In the present work it is suggested that for fitting natural aerosol size distributions, the use of Chebyshev polynomials is highly favorable. In this case, formula 4.2 is called a Chebyshev expansion and is of the form

$$F(t) = b_o T_o + b_1 T_1 + \ldots + b_k T_k. \tag{4.8}$$

This saves computing time considerably, and enables an easy check for convergence. However, it is then more difficult to interpret the results of the fit in a straight forward and general way.

Chebyshev polynomials are well known from many textbooks on approximation theory and smoothing. There, one usually may also find an outline of their special properties which makes them a rather unique set of basis functions for use in numerical fit algorithms. The first few members of this set are given by

$$T_o(t): = 1$$

$$T_1(t): = t$$

$$T_2(t): = -1 \quad +2t^2$$

$$T_3(t): = -3t \quad +4t^3$$

$$T_4(t): = 1 \quad -8t^2 \quad + 8t^4$$

$$T_5(t): = 5t \quad -20t^3 +16t^5. \tag{4.9}$$

The first polynomials T_o to T_7 are shown graphically in Figure 2 for the interval $-1 \le t \le 1$. From this figure, it may be seen that

$$\left| T_k (t) \right| \le 1 \quad ; \quad \left| T \right| \le 1. \tag{4.10}$$

9

The definitions (4.9) indicate that the polynomials T_k are built up from either even or odd power functions in t. The amplitudes of the power functions obviously tend to be vary large in magnitude, and it is because of their alternate signs that the T_k stay within the limits (4.10). The numerical evaluation of series like this may generally be subject to serious computational errors for computers having limited word lengths. A convenient way out of this problem is to use the three-terms recurrence relation

$$T_{k+2}(t) = 2 t \cdot T_{k+1}(t) - T_k(t) \quad , \quad k=0,1,2,\ldots \qquad (4.11)$$

in conjunction with the basic definition (4.9) for T_o and T_1.

Some Examples

Though the examples chosen are rather simple, they are of general interest in the study of aerosols. They are:

- The Junge size distribution
- A unimodal lognormal size distribution
- The haze L model

The Junge size distribution is of the form

$$n(D) := \frac{dN}{dC} = C \, D^{-(\beta+1)} \, , \qquad (4.12)$$

where the prime is to indicate that, since n is a function of diameter rather than of radius, the concentration C differs in value from those to be found in conventional tabulations.

On a log-log scale, n(D) plots as a straight line. In this case, the Chebyshev fit results in a linear function

$$F(t) = \sum_{k=0}^{1} b_k T_k(t)$$

$$= b_o + b_1 t. \qquad (4.13)$$

10

The interpretation of the two amplitudes, b_o and b_1, is most conveniently done in the double logarithmic plot. It may be seen that b_o mainly determines the absolute concentration, while b_1 measures the slope.

Junge distribution and Chebyshev approximations

The Junge parameters, C' and B can be expressed uniquely in terms of the amplitudes, b_0 and b_1.

$$\log C' = b_0 - b_1 \left(\frac{\log (D_H \cdot D_L)}{\log (D_H / D_L)} \right) \qquad (4.14)$$

$$B + 1 = \frac{2}{\log (D_H / D_L)} \cdot b_1 \cdot$$

From this it is clear that the amplitudes b_k are dependent on the low and high ends, D_L and D_H respectively, of the measured diameter interval. Thus, when compairing Chebyshev amplitudes for measurements taken with different devices, the corresponding values for D_L and D_H must be taken into consideration.

In case of Junge distributions, we can avoid this unpleasant device-dependence. However, the problem is much more involved for more general distributions containing many more terms. We hope to resolve this problem in future studies.

The next example is that of a unimodal lognormal distribution

$$n(D) = \frac{2N'}{D. \log \sigma \sqrt{2 \pi}} \exp \left\{ - \frac{1}{2} \left(\frac{\log \frac{D}{D_m}}{\log \sigma g} \right) \right\}$$

Again, the prime indicates that in the current notation the parameter N is different in value from that in the literature where n has been considered to be a function of the particle radius. The example illustrated is an average size distribution found by Jaenicke et al. for maritime aerosols, and

12

taken here at the experimental meshpoints of the present work. The median number diameter accordingly is $D_m = 0.52 \mu m$; the logarithmic standard deviation is $\sigma_g = 4.8$.

One can see that the Chebyshev fit of a unimodel lognormal distribution is excellent. Moreover, three coefficients are sufficient to specify this fit, i.e. b_o, b_1' and b_2 as one would expect since a lognormal distribution is represented in the double logarithmic scale by a parabola of order 2.

The physical parameters N', D_m and b_b can be expressed uniquely in terms of the amplitudes b_o' b_1' and b_2.

The last example we consider is that of Deirmendjian's Haze L model distribution

$$n (D) = a \quad D \quad \exp \quad (-b \cdot D^{\delta}) \qquad (4.15)$$

with the following parameter values

$$a' = 2.4878 \quad 10^6 \, [\mu^{- -1}] \,, \quad \delta = 2$$

$$b' = 10.6905 \, [\mu^{\gamma}]_1 \,, \quad \epsilon = 5$$

Unlike the two other examples in the log-log scale the distribution does not represent a simple polynomial. Thus, the Chebyshev fit in this case is really an approximation.

The few examples presented have been chosen to demonstrate the main advantages and shortcomings of fit time algorithms. The fit was shown to be versatile in the sense that various size distributions can be treated although they are of quite different nature. All of them are basically related to the aerosol field. Moreover, unlike in cases of orthogonal polynomials, the fixed basis of Chebyshev polynomials for the function space permits a straight-

13

forward, transparent interpretation of the fitted data in terms of a reduced data set, namely the resulting Chebyshev amplitudes. A remaining difficulty lies in the fact that the resulting amplitudes are dependent on the range of particle diameters covered in the fit, and therefore to some extent are device-dependent.